PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-229985
(43)Date of publication of application : 16.08.2002

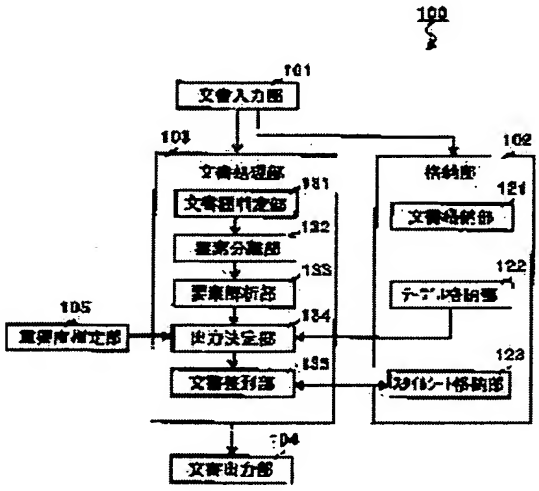
(51)Int.Cl. G06F 17/30
G06F 12/00
G06F 17/21

(21)Application number : 2001-030260
(22)Date of filing : 06.02.2001

(71)Applicant : RICOH CO LTD
(72)Inventor : NARAHARA KOICHI

(54) APPARATUS AND METHOD FOR STRUCTURED DOCUMENT PROCESSING, AND PROGRAM FOR MAKING COMPUTER EXECUTE THE STRUCTURED DOCUMENT PROCESSING

(57)Abstract:
PROBLEM TO BE SOLVED: To enhance the accuracy of judgment at outputting of important elements.
SOLUTION: A PC100 applying the apparatus for structured document processing to a personal computer is an apparatus for structured document processing to make the contents of the elements described in a structured document description language able to be outputted and provided with a document input part 101 inputting a structured document described in the structured document description language, an output decision part 134 deciding whether the elements should be outputted or not according to the kinds of tags of each element in the structured document inputted from the input part 101 and a document shaping part 135 processing the elements decided to be outputted by the decision part 134 while making them able to be outputted.



LEGAL STATUS

[Date of request for examination] 21.10.2004

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2002－229985
（P2002－229985A）
(43)公開日　平成14年８月16日(2002.8.16)

(51)Int.Cl.⁷		識別記号	FI		テーマコード（参考）
Ｇ０６Ｆ	17/30	１４０	Ｇ０６Ｆ 17/30	１４０	５Ｂ００９
		３８０		３８０Ｚ	５Ｂ０７５
	12/00	５４７	12/00	５４７Ｈ	５Ｂ０８２
	17/21	５０１	17/21	５０１Ｔ	

審査請求　未請求　請求項の数10　ＯＬ　（全 13 頁）

(21)出願番号　特願2001－30260(P2001－30260)

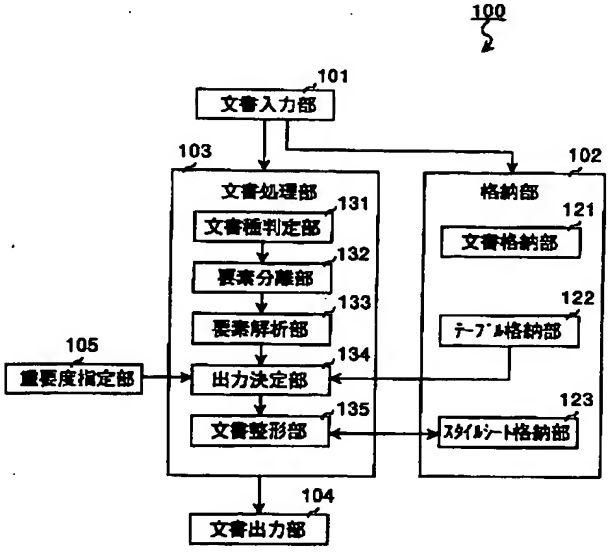
(22)出願日　平成13年２月６日(2001.2.6)

(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(72)発明者　楢原　孝一
東京都大田区中馬込１丁目３番６号　株式会社リコー内
(74)代理人　100089118
弁理士　酒井　宏明

Ｆターム(参考)　5B009 NA05
5B075 ND26 NR02 NR20
5B082 GA08

(54)【発明の名称】　構造化文書処理装置、構造化文書処理方法およびコンピュータに構造化文書処理を実行させるためのプログラム

(57)【要約】
【課題】　重要な要素を出力させる際の判定精度を高めること。
【解決手段】　構造化文書処理装置をパーソナルコンピュータに適用したＰＣ１００は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、構造化文書記述言語により記述された構造化文書を入力する文書入力部１０１と、文書入力部１０１により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定部１３４と、出力決定部１３４により出力させると決定された要素を出力可能に処理する文書整形部１３５と、を備える。



1

【特許請求の範囲】

【請求項1】　構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、
前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、
前記構造化文書入力手段により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定手段と、
前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段と、
を備えたことを特徴とする構造化文書処理装置。

【請求項2】　構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、
前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、
前記構造化文書入力手段により入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定手段と、
前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段と、
を備えたことを特徴とする構造化文書処理装置。

【請求項3】　前記構造化文書処理手段により出力可能に処理された前記要素の内容を出力する構造化文書出力手段を具備したことを特徴とする請求項1または2に記載の構造化文書処理装置。

【請求項4】　前記構造化文書出力手段は、表示装置もしくは印刷装置であることを特徴とする請求項3に記載の構造化文書処理装置。

【請求項5】　構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理方法であって、
前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力工程と、
前記構造化文書入力工程で入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定工程と、
前記出力決定工程で出力させると決定された要素を出力可能に処理する構造化文書処理工程と、
前記構造化文書処理工程で出力可能に処理された前記要素の内容を出力する構造化文書出力工程と、
を含んだことを特徴とする構造化文書処理方法。

【請求項6】　構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理方法であって、
前記構造化文書記述言語により記述された構造化文書を

2

入力する構造化文書入力工程と、
前記構造化文書入力工程で入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定工程と、
前記出力決定工程で出力させると決定された要素を出力可能に処理する構造化文書処理工程と、
前記構造化文書処理工程で出力可能に処理された前記要素の内容を出力する構造化文書出力工程と、
を含んだことを特徴とする構造化文書処理方法。

【請求項7】　前記構造化文書出力工程では、表示装置もしくは印刷装置を介して前記要素の内容を出力することを特徴とする請求項5または6に記載の構造化文書処理方法。

【請求項8】　構造化文書記述言語により記述された要素の内容を出力可能に処理するプログラムであって、
コンピュータを、
前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、
前記構造化文書入力手段により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定手段と、
前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段として機能させることを特徴とするプログラム。

【請求項9】　構造化文書記述言語により記述された要素の内容を出力可能に処理するプログラムであって、
コンピュータを、
前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、
前記構造化文書入力手段により入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定手段と、
前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段、
として機能させることを特徴とするプログラム。

【請求項10】　さらに、コンピュータを、前記構造化文書処理手段により出力可能に処理された前記要素の内容を出力する構造化文書出力手段として機能させるプログラムを含んだことを特徴とする請求項8または9に記載のプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、構造化文書処理装置、構造化文書処理方法、コンピュータに構造化文書処理を実行させるためのプログラムに関し、特に、構造化

文書記述言語で記述された文書を、タグの種類や個数に基づいて表示または印刷する構造化文書処理装置、構造化文書処理方法および、コンピュータに構造化文書処理を実行させるためのプログラムに関する。

【０００２】

【従来の技術】従来、インターネット環境の発展によりＨＴＭＬ（ＨｙｐｅｒＴｅｘｔ　Ｍａｒｋｕｐ　Ｌａｎｇｕａｇｅ）、ＸＭＬ（ｅＸｔｅｎｓｉｂｌｅ　ＭａｒｋｕｐＬａｎｇｕａｇｅ：拡張可能な印付け言語）などに代表される構造化文書記述言語が広く利用されている。

【０００３】構造化文書記述言語とは、構造化文書を記述するための規約である。構造化文書は、要素の集合からなり、各要素は、タグと要素の内容とから構成される。要素の内容とは、構造化文書の作成者が表示させたいと考える文書や図形などの構造化文書の実体的部分をいう。タグとは、その要素の内容を表示する際のフォントの大きさなど、その要素の出力態様ないし属性を指定する構造化文書の規約的部分をいう。

【０００４】構造化文書の作成者は、出力させたい要素の内容に、タグという印付けをおこない、構造化文書を作成する。なお、要素の内容は、文字データや画像データの他にも、音声データなどを含めることも可能である。

【０００５】ＨＴＭＬとＸＭＬの違いは利用可能なタグの種類にある。ＨＴＭＬはあらかじめ定められた約８０種類のタグを使用する言語であるのに対し、ＸＭＬは文書作成者が自由にタグの種類を設計可能な言語である。

【０００６】図１１は、ＨＴＭＬで記述した文書情報、いわゆるソースであり、図１２は、ＸＭＬで記述した文書情報（ソース）である。これらを、Ｉｎｔｅｒｎｅｔ　Ｅｘｐｌｏｒｌｅｒ（マイクロソフト社の登録商標）やＮｅｔｓｃａｐｅ　Ｎａｖｉｇａｔｏｒ（Ｎｅｔｓｃａｐｅ　Ｃｏｍｍｕｎｉｃａｔｉｏｎｓ社の登録商標）といったブラウザで処理すると、図１３に示した内容で出力、すなわち、コンピュータ画面上に表示される。また、一定の操作をおこなうことにより、プリンタから出力、すなわち、印刷することもできる。

【０００７】図１３に示したように、図１１もしくは図１２に示したソースからはいずれも同一の出力が得られる。一方、図１１および図１２に示したように、同一の出力結果を得るソースであっても、ＨＴＭＬとＸＭＬでは、使用されているタグの種類が異なっていることが確認できる。ＸＭＬは、ＨＴＭＬと比較して、要素の内容の出力態様ないし属性を詳細に記述できる点が大きく異なる。従来では、構造化文書記述言語により、豊富なコンテンツを閲覧者ないし利用者に提供することが可能であった。

【０００８】

【発明が解決しようとする課題】しかしながら、従来では以下のような問題点があった。図１３の例では、出力すべき内容が少ないので、ＣＲＴなどの画面に表示する場合は画面内に文書全体が表示可能であり、印刷装置などで印刷する場合はＡ４サイズの用紙1枚以内に印刷可能である。しかしながら、出力すべき内容は一画面分もしくは1ページ分に限られるわけではないので、この場合は画面のスクロールが必要であったり、複数ページに印刷する必要がある。

【０００９】このとき、その出力内容を見る者の閲覧効率や利用効率のため、重要な要素のみを選択して表示または印刷する技術が求められていた。たとえば、従来では、文字データ、画像データ、ＵＲＬ参照ポインタのといった各要素の内容自体の種別や各要素の内容のデータ量にしたがって、その要素が重要であるか否かを判断していた。換言すると、要素が重要であるか否かを、要素の物理的特性によって判断していた。

【００１０】また、特開平１１―２０３１００号「ネットワークプリンタ及びネットワーク印刷方法」では、ＨＴＭＬで記述された文書情報から重要な要素を利用者側が判断し選択する技術が開示されている。しかしながら、この場合でもやはり、選択の基準が、文字データのみを持つ要素であったり、画像データのみを持つ要素であったり、また、データ量や画像サイズが少ない要素を対象としている。したがって、この従来技術も、要素が重要であるか否かを、要素の物理量によって判断したものであった。

【００１１】このような判断をおこなうと、要素の意味内容に即して重要度が判断されるわけではないので、次のような誤判定が生じるという問題点があった。例えば、長い文章で構成される要素と短い文章で構成される要素とがあった場合に、従来技術ではデータ量の少ない文章を重要なデータとして判定している。しかしながら、長い文章の方が重要な場合もあり、この場合は誤判定になるという問題点があった。

【００１２】また、ＷＷＷ（Ｗｏｒｌｄ　Ｗｉｄｅ　Ｗｅｂ）等で提供されている文書情報には、広告等の重要でない冗長な要素が含まれている場合がある。従来技術では、ＵＲＬ参照ポインタと画像データから構成されている要素を、広告情報であり、重要でないと判断しているが、重要な要素が広告と同様の構成で記述されている場合もあり、この場合は誤判定になるという問題点があった。

【００１３】すなわち、従来の技術では、重要な要素を出力させる場合に、要素の物理量にしたがって当該要素を重要か否かを判断していたので、判定精度が必ずしも高くないという問題点があった。

【００１４】本発明は上記に鑑みてなされたものであって、重要な要素を出力させる際の判定精度を高めることを目的とする。

【００１５】

5

【課題を解決するための手段】上記の目的を達成するために、請求項1に記載の構造化文書処理装置は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、前記構造化文書入力手段により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定手段と、前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段と、を備えたことを特徴とする。

【００１６】すなわち、請求項1にかかる発明は、タグを基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定できる。

【００１７】また、請求項2に記載の構造化文書処理装置は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、前記構造化文書入力手段により入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定手段と、前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段と、を備えたことを特徴とする。

【００１８】すなわち、請求項2にかかる発明は、タグと各要素間の論理構造を基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定できる。

【００１９】また、請求項3に記載の構造化文書処理装置は、請求項1または2に記載の構造化文書処理装置において、前記構造化文書処理手段により出力可能に処理された前記要素の内容を出力する構造化文書出力手段を具備したことを特徴とする。

【００２０】また、請求項3にかかる発明は、構造化文書のうちの重要な要素の内容を出力する。

【００２１】また、請求項4に記載の構造化文書処理装置は、請求項3に記載の構造化文書処理装置において、前記構造化文書出力手段が、表示装置もしくは印刷装置であることを特徴とする。

【００２２】すなわち、請求項4にかかる発明は、重要な要素を出力させる際の判定精度を高める構造化文書処理装置を提供することができる。

【００２３】また、請求項5に記載の構造化文書処理方法は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理方法であって、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力工程と、前記構造化文書入力工程で入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定工程と、前記出力決定工程で出力させると決定された要素を出力可能に処理する構造化文書処理工程と、前記構造化文書処理工程で出力可能に処理された前記要素の内容を出力する構造化文書出力工程と、を含んだことを特徴とする。

6

【００２４】すなわち、請求項5にかかる発明は、タグを基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定できる。

【００２５】また、請求項6に記載の構造化文書処理方法は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理方法であって、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力工程と、前記構造化文書入力工程で入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定工程と、前記出力決定工程で出力させると決定された要素を出力可能に処理する構造化文書処理工程と、前記構造化文書処理工程で出力可能に処理された前記要素の内容を出力する構造化文書出力工程と、を含んだことを特徴とする。

【００２６】すなわち、請求項6にかかる発明は、タグと各要素間の論理構造を基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定できる。

【００２７】また、請求項7に記載の構造化文書処理方法は、請求項5または6に記載の構造化文書処理方法において、前記構造化文書出力工程では、表示装置もしくは印刷装置を介して前記要素の内容を出力することを特徴とする。

【００２８】すなわち、請求項7にかかる発明は、重要な要素を出力させる際の判定精度を高める構造化文書処理方法を提供することができる。

【００２９】また、請求項8に記載のプログラムは、構造化文書記述言語により記述された要素の内容を出力可能に処理するプログラムであって、コンピュータを、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、前記構造化文書入力手段により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定手段と、前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段として機能させることを特徴とする。

【００３０】すなわち、請求項8にかかる発明は、タグを基に重要度を判定させ、構造化文書内で重要な意味を持つ要素を特定させることができる。

【００３１】また、請求項9に記載のプログラムは、構造化文書記述言語により記述された要素の内容を出力可能に処理するプログラムであって、コンピュータを、前記構造化文書記述言語により記述された構造化文書を入

力する構造化文書入力手段と、前記構造化文書入力手段により入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定手段と、前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段、として機能させることを特徴とする。

【００３２】すなわち、請求項9にかかる発明は、タグと各要素間の論理構造を基に重要度を判定させ、構造化文書内で重要な意味を持つ要素を特定させることができる。

【００３３】また、請求項１０に記載のプログラムは、請求項8または9に記載のプログラムにおいて、さらに、コンピュータを、前記構造化文書処理手段により出力可能に処理された前記要素の内容を出力する構造化文書出力手段として機能させるプログラムを含んだことを特徴とする。

【００３４】すなわち、請求項１０にかかる発明は、構造化文書の重要な要素の内容を出力させる。

【００３５】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

実施の形態1．実施の形態1では、構造化文書記述言語としてＸＭＬが使用された構造化文書を入力し、要素のタグないし論理構造にしたがって、その内容を出力する構造化文書処理装置を、パーソナルコンピュータ（ＰＣ）に適用した例について説明する。ここでは、まず、ＸＭＬが使用された構造化文書について説明し、つぎに装置（ＰＣ）について説明する。

【００３６】（ＸＭＬが使用された構造化文書の説明）図1は、実施の形態1で使用される、ＸＭＬが使用された構造化文書の構成例を示した図である。なお、図1は、説明の便宜上、図１２で示した構造化文書と同一の構成としている。また、図1には便宜的に行番号を記しているが実際に必要とされるものではない。

【００３７】構造化文書とは、前述したように、構造化文書記述言語で記述された文書であり、要素の内容、すなわち、文書作成者が伝達したい情報（文章、画像、音声など）が、山パーレン＜＞で示されるタグで囲まれた文章である（図1参照）。タグは要素の内容を表示する際のフォントの大きさなど、その要素の出力態様ないし属性を指定する構造化文書の規約的部分である。図1に示した例でいうと、ｄｏｃ、ｔｉｔｌｅ、ｍａｉｎ等の文字列がタグである。

【００３８】構造化文書は、ＸＭＬ宣言、文書型宣言、文書エンティティの3つのブロックから構成される。図1の例では、1行目がＸＭＬ宣言、3～１２行目が文書型宣言、１４～３１行目が文書の実体的部分、すなわち、文書作成者が閲覧者ないし利用者に伝達したい内容をしるす部分である。なお、この文書の実体的部分を、以降において適宜文書エンティティと称する。

【００３９】つぎに、構造化文書の各ブロックについて説明する。ＸＭＬ宣言とは、この構造化文書がＸＭＬで記述されていることを明示する宣言である。文書型宣言とは、構造化文書に含まれる要素の属性や文書の論理構造といった文書型定義がなされる宣言である。たとえば、文書型定義では要素の名前、要素間の親子関係、子要素の出現順序、出現回数などを規定する。

【００４０】論理構造とは、タグの入籠構造をいう。図２は、構造化文書のタグの入籠構造の一例を示した説明図である。図には、図1に示した構造化文書のタグの入籠構造を示している。図に示したように、タグ「ｄｏｃ」はルート（最上位の要素の属性を示すタグ）に相当し、下位に、「ｔｉｔｌｅ」、「ｍａｉｎ」、「ｍｉｓｃ」、「ｉｍｇ」を含んだ構造となっている。

【００４１】文書エンティティは出力させたい文書の実体的内容を記述したブロックである。この文書エンティティは、ルートとなる要素中で、始まりを示すタグ（開始タグ）で始まり、終わりを示すタグ（終了タグ）で終わる。全ての要素は開始タグ、終了タグを持ち、各要素の内容は開始タグと終了タグの間に記述する。開始タグは＜タグ名＞、終了タグは＜／タグ名＞と記述し、タグ名には文書型定義で定義した要素の名前を用いる。たとえばタグ「ｄｏｃ」の開始タグは「＜ｄｏｃ＞」、終了タグは「＜／ｄｏｃ＞」である。

【００４２】以上説明したように、ＸＭＬを用いて記述された構造化文書により、後述する構造化処理装置を用いて、重要度に応じて出力スタイルを変更することが可能となる。

【００４３】（構造化文書処理装置の内容）つぎに、本願発明の構造化文書処理装置をパーソナルコンピュータ（ＰＣ）に適用した例を図面を参照しながら説明する。図3は、本発明を実施する構造化文書処理装置をＰＣに適用した例の機能ブロック図である。図4は、本発明を実施するＰＣの構成例を示した説明図である。

【００４４】ＰＣ１００は、構造化文書を入力する文書入力部１０１と、文書入力部１０１で入力された構造化文書を初め後述するテーブルやスタイルシートを格納する格納部１０２と、構造化文書を出力可能に処理する文書処理部１０３と、文書処理部１０３で処理された構造化文書を実際に出力する文書出力部１０４と、どの重要度まで出力させるかを指定する重要度指定部１０５と、を有する。なお、出力とは、表示、印刷、スピーカからの音声出力など、人間の五官により知覚可能に処理されたものすべてを含む。

【００４５】ハードウェア構成としては、ＰＣ１００は、構造化文書を出力可能に変換処理するＣＰＵ２０１（図4参照）と、ＣＰＵ２０１のワークエリアであるＲＡＭ２０２と、ＯＳを含み様々なソフトウェアを格納

し、また、構造化文書を含み様々なファイルを格納するハードディスク２０３と、構造化文書を表示するＣＲＴ２０４と、ＣＲＴ２０４の出力制御をおこなうビデオカード２０５と、構造化文書を印刷するプリンタ２０６と、各種の指示をおこなうキーボード２０７と、プリンタ２０６やキーボード２０７の入出力を制御するＩ／Ｆ２０８と、インターネットに接続しＷｅｂサーバから構造化文書ファイルを入力するモデム２０９と、を有する。

【００４６】また、ハードディスク２０３は、ＰＣ１００の基本動作を制御するＯＳ２３１と、構造化文書を解析し、出力可能に処理するプログラムであるＸＭＬ解析アプリケーション２３２と、プリンタドライバ２３３と、ブラウザ２３４と、を有する。なお、使用の態様によっては、ＸＭＬ解析アプリケーション２３２は、ブラウザ２３４やＯＳ２３１に組み込まれていてもよい。また、ブラウザ２３４はＯＳ２３１に組み込まれていてもよい。

【００４７】ハードディスク２０３は、さらに、図1に示したような構造化文書を電子ファイルである構造化文書ファイル２３５として格納する。ハードディスク２０３は、この他、構造化文書ファイル２３５にリンクの張られている画像ファイル２３６、画像ファイル２３７および音声ファイル２３８等を格納する。

【００４８】また、ハードディスク２０３は、文書処理部１０３で使用するテーブル２３９と、スタイルシート２４０とを格納する。なお、後に詳述するが、テーブル２３９は、要素の内容を出力させるか否かを決定する際の判断材料として用いられる参照テーブルであり、スタイルシート２４０は、出力フォーマットを決定する補助情報である。

【００４９】つぎに、各部の内容を説明する。

（文書入力部１０１の内容）文書入力部１０１は、構造化文書を入力する。入力とは、構造化文書をエディタを用いて入力することを意味する場合もあれば、インターネット等を通じてＷｅｂサーバから入力することも意味する。また、ハードディスクに２０３に格納された構造化文書ファイル２３５を読み出すことであってもよい。すなわち、入力とは、ＰＣ１００内に構造化文書が作成ないし取り込まれることを意味する。文書入力部１０１は、モデム２０９とＯＳ２３１もしくはキーボード２０７によりその機能を実現することができる。

【００５０】（格納部１０２の内容）格納部１０２は、文書入力部１０１で入力された構造化文書を格納する文書格納部１２１と、後述する出力決定部で参照するテーブル２３９を格納するテーブル格納部１２２と、後述する文書整形部で使用するスタイルシートを格納するスタイルシート格納部１２３とを有する。格納部１０２は、ハードディスク２０３およびＲＡＭ２０２によりその機能を実現することができる。

【００５１】また、ハードディスク２０３やＲＡＭ２０２以外でも、ＣＤ－ＲＯＭ、ＭＯなどによりその機能を実現することができる。なお、構造化文書は、テキスト文書を入力するいわゆるエディタを用いて作成することができる。また、専用のエディット機能を持つエディタを利用し、文書型定義で規定された論理構造にしたがって文書を作成することもできる。

【００５２】（文書処理部１０３の内容）文書処理部１０３は、構造化文書をＣＲＴ２０４やプリンタ２０６で出力可能に処理する。文書処理部１０３は、入力した文書がＸＭＬの使用された構造化文書であるかを判定する文書種判定部１３１と、ＸＭＬが使用された構造化文書の要素をタグと要素の内容とに分離する要素分離部１３２と、要素の構造を後述するツリー構造のデータとして解析する要素解析部１３３と、ツリー構造のデータを参照しつつ、タグやタグの入籠構造もしくは要素の内容の重要度を入力し、どの要素の内容を出力すべきか決定する出力決定部１３４と、決定された要素の内容を整形する文書整形部１３５と、を有する。

【００５３】文書処理部１０３は、ＯＳ２３１、ＸＭＬ解析アプリケーション２３２、プリンタドライバ２３３、ブラウザ２３４、ＣＰＵ２０１によりその機能を実現することができる。なお、文書処理部１０３の具体的な処理内容については後述する。

【００５４】（文書出力部１０４の内容）文書出力部１０４は、文書処理部１０３で処理された文書を出力する。具体的には、ＣＲＴ２０４やプリンタ２０６から構成される。なお、スピーカも含まれる。これは、構造化文書で音声ファイルが参照されている場合には、スピーカからその音声ファイルの内容が出力されるからである。

【００５５】つぎに、文書処理部１０３の処理内容を詳述する。

（文書処理部１０３：文書種判定部１３１の内容）文書種判定部１３１は、読み込んだファイルがＸＭＬで記述されているかを判定する。ＸＭＬが使用された構造化文書の場合には、上述したように、ＸＭＬ宣言、文書型宣言、文書エンティティの3ブロックが順に記載されている。したがって、文書種判定部１３１は、ファイルの内容を順次読み込み、ＸＭＬ宣言ブロックが記述されているかを判定する。文書種判定部１３１は、ＸＭＬ解析アプリケーション２３２もしくはブラウザ２３４、およびＣＰＵ２０１によりその機能を実現することができる。

【００５６】（文書処理部１０３：要素分離部１３２の内容）要素分離部１３２は、文書種判定部１３１によりＸＭＬが使用された構造化文書であると判定されたファイルの各要素をタグと要素の内容とに分離する。この分離により構造化文書のタグの構造を管理する要素解析部１３３での処理が容易となる。要素分離部１３２は、ＸＭＬ解析アプリケーション２３２およびＣＰＵ２０１に

よりその機能を実現することができる。

【００５７】（文書処理部１０３：要素解析部１３３の内容）要素解析部１３３は、要素分離部１３２で分離されたタグと要素の内容を構文解析ツリー（以降では適宜構文解析木と称する）と呼ばれる図2に示したようなツリー構造のデータに振り分け管理する。すなわち、木構造の各節を、タグ、属性、要素の内容を一組として管理する。なお、図2では要素の内容の表示を省略している。図に示したように、タグ「ｄｏｃ」の中には、さらにタグ「ｍａｉｎ」が定義され、このタグ「ｍａｉｎ」では、さらにタグ「ｓｅｃｔｉｏｎ」が定義されている。

【００５８】すなわち、タグが階層的に定義されている。なお、ここで説明したタグ「ｍａｉｎ」は、図から明らかなように、さらに2階層の深さのタグを有している。要素解析部１３３は、ＸＭＬ解析アプリケーション２３２と、ＣＰＵ２０１によりその機能を実現することができる。

【００５９】（文書処理部１０３：出力決定部１３４の内容）出力決定部１３４は、テーブル格納部１２２に格納されているテーブル２３９を参照して、タグに使われている文字列の意味を解析し、文書出力部１０４から出力させるべき要素であるか否かを決定する。出力決定部１３４は、出力決定の指標とすべくタグの内容を、まずカテゴリに分類する。

【００６０】ここで、ＸＭＬではタグは自由に設計することが可能なため、使用するタグには要素の意味を表す文字列を用いることができる。実際に、文章のタイトルを表す要素には「ｔｉｔｌｅ」、「Ｔｉｔｌｅ」、「タイトル」等のタグを用い、文書の内容を表す要素には「ｃｏｎｔｅｎｔｓ」、「Ｃｏｎｔｅｎｔｓ」、「内容」、「本文」等のタグを用いることができる。

【００６１】図5は、テーブル格納部１２２に格納されたテーブル２３９の内容の一例を表した説明図である。テーブル２３９の左列はタグの意味を表すカテゴリ、中央列はカテゴリに属するタグ、右列は要素内容を出力させるか否かを決定する重要度であり、各カテゴリに対して付与されている。なお、図の例ではカテゴリと重要度が1対1に対応しているが、これに限られるものではない。

【００６２】図6は、カテゴリと、タグと、重要度との関係の他の例を示した図である。Ｃ１は、タイトルに関連するタグのカテゴリであるが、タグが日本語で表示されるもの（具体的にはタグ「タイトル」）については重要度がｌｅｖｅｌ１、タグが英語で表示されるもの（具体的にはタグ「ｔｉｔｌｅ」、「Ｔｉｔｌｅ」）については重要度がｌｅｖｅｌ２に設定されている。この様に分類しておくことで、文書作成者は、日本語タグに対しては、重要度のより高い要素の内容を記述し、英語タグに対しては、重要度の少し低い内容を記述する等してタグを使い分けることができる。たとえば言語の異なる国に同一コンテンツを配信する際に役立つ。

【００６３】テーブル２３９を参照することにより、木構造の各節に対する意味解析処理が行われる。本実施例ではタグの文字列を解析することにより意味を推定しているが、この方式に限定することはなく他の方式を用いても構わない。意味解析処理は、図2に示した構文解析木の各節で管理されているタグないし要素の内容の重要度を判定する処理である。重要度の判定は重要度に関する情報が登録されたテーブル２３９を利用する。図5に示したテーブル２３９のｌｅｖｅｌ１、ｌｅｖｅｌ２、ｌｅｖｅｌ３がカテゴリに対応した重要度であり、数値が小さいほど重要度が高い情報であることを示す。出力決定部１３４は、このテーブル２３９を参照することにより図2に示した構文解析木の各節の重要度を判定する。判定結果は構文解析木と共に格納部１０２に記憶してもよい。

【００６４】重要度を判定する際には、ユーザが重要度を指定し、指定された重要度よりも高い重要度の要素を出力可能に処理する。この指定は、重要度指定部１０５によりおこなう。重要度指定部１０５は、キーボード２０７によりその機能を実現することができるが、この他、マウスやバーコードリーダなどによって入力してもよい。たとえば、ユーザが重要度としてｌｅｖｅｌ２を指定した場合は、ｌｅｖｅｌ１、ｌｅｖｅｌ２に対応するタグの要素が全て選択され、出力可能に処理される。

【００６５】なお、使用の態様によっては、重要度でなくカテゴリを指定して、指定されたカテゴリを出力するようにしてもよい。出力決定部１３４は、ＸＭＬ解析アプリケーション２３２、テーブル２３９、スタイルシート２４０、ＣＰＵ２０１によりその機能を実現することができる。なお、ここでは、テーブル格納部１２２に格納したテーブル２３９を出力決定部１３４が参照する例について説明したが、これに限ることなく、要素解析部１３３が参照する態様であってもよい。

【００６６】（文書処理部１０３：文書整形部１３５の内容）文書整形部整形装置は、重要項目選択装置で選択された要素に対して文書整形の規則を定めたスタイルシート２４０にそって整形処理を実施する。図7は、ＣＳＳ（Ｃａｓｃａｄｉｎｇ　Ｓｔｙｌｅ　Ｓｈｅｅｔｓ）と呼ばれるスタイルシートの一例を示した図である。スタイルシートは、構造化文書を構成する各要素の内容を出力する際の文字サイズ、フォントの種類などを指定するシートである。

【００６７】指定は、各タグに対しておこなう。図7の例では、タグ「ｔｉｔｌｅ」は２４ポイントのフォントで太字に指定するものであり、タグ「ｓｅｃｔｉｏｎｔｉｔｌｅ」は１８ポイントのフォントで太字、斜体に設定するものである。他の要素についても必要に応じて指定することが可能であるがここでは省略する。

13

【００６８】以上のようなスタイルシートにより画面での表示、紙への印刷のための整形処理が文書処理部１０３で実施され、整形された構造化文書は文書出力部１０４から出力される。出力例を図8に示す。ここでは、ｌｅｖｅｌ１およびｌｅｖｅｌ２に対応するタグである「ｔｉｔｌｅ」、「ｓｅｃｔｉｏｎｔｉｔｌｅ」の要素から構成される文書が出力された例を示している。

【００６９】文書整形部１３５は、ＸＭＬ解析アプリケーション２３２、スタイルシート２４０、ＯＳ２３１およびＣＰＵ２０１によりその機能を実現することができる。

【００７０】（構造化文書の処理の流れ）つぎに、ＰＣ１００の具体的な処理の流れを説明する。図9は、構造化文書の処理の流れの一例を示したフローチャートである。文書入力部101は、電子ファイルを入力データとして読み込む（ステップＳ９０１）。電子ファイルは、モデム２０９からインターネット経由で読み込んでもよいし、場合によっては、既にハードディスク２０３に格納されているものを読み出してもよい。

【００７１】つぎに、文書種判定部１３１は、ファイルの先頭を読み出し、入力した文書がＸＭＬ文書ファイルであるか否かを判定する（ステップＳ９０２）。ＸＭＬ文書でないときは（ステップＳ９０２：ＮＯ）、処理を終了し、ＸＭＬ文書であるときは（ステップＳ９０２：ＹＥＳ）、文書型宣言と文書エンティティを読み込む（ステップＳ９０３）。要素分離部１３２および要素解析部１３３は、この読み込まれた文書型宣言と文書エンティティとを、タグと要素の内容とに分離し、これらを構文解析木として管理する（ステップＳ９０４）。

【００７２】出力決定部１３４は、構文解析木により管理される要素のうち、出力すべき要素を、テーブル２３９を参照することにより決定する（ステップＳ９０５）。つづいて、文書整形部１３５は、スタイルシート２４０を読み込み、タグにしたがって、要素の内容を処理し、文書を整形する（ステップＳ９０６）。なお、この整形は、フォントの大きさやフォントの飾り（斜体、太字、下線）の他、音声ファイルの場合は、出力するボリュームの大きさや各種のサウンドエフェクトを施す。また、動画の場合はその大きさや使用する色を調整する（たとえば、カラー画像を白黒画像やセピア色に調整する）。

【００７３】最後に、ＣＲＴ２０４や図示しないスピーカは処理された要素の内容を出力（表示）する（ステップＳ９０７）。なお、ここではＣＲＴ２０４からの表示を述べたが、プリンタ２０６から出力（印刷）してもよい。

【００７４】以上説明したように、実施の形態1のＰＣは、構造化文書のタグの種類にしたがって要素の内容を出力するので、要素の物理量といった画一的な判断によらず、要素の意味内容を反映した出力が可能となる。また、構造化文書作成者は、伝達したい要素の内容にかかるタグを重要度の高いものとしてテーブル化することができ、意図した内容を閲覧者ないし利用者に伝達することができる。

14

【００７５】実施の形態2．実施の形態2では、構造解析木に基づいて要素の内容を出力する構造化文書処理装置をＰＣに適用した例について説明する。なお、実施の形態2では、実施の形態1の構成部分と同一の構成部分については、その説明を省略し、特に断らない限り、同一の符号を付することとする。

【００７６】実施の形態2では、要素解析部１３３がテーブル格納部１２２に格納されたテーブル２３９を参照する。すなわち、要素解析部１３３は、要素分離部１３２で分離されたタグと要素の内容を構文解析木に振り分けて管理する。すなわち、木構造の各節を、タグ、属性、要素の内容に加えて、テーブル２３９を参照することにより各要素の重要度とカテゴリーも一組として管理する。また、この構文解析木を構築することにより、各要素の階層、すなわち、木構造の深さも管理されることとなる。

【００７７】図１０は、実施の形態2の構造化文書処理装置をＰＣに適用した場合の構造化文書の処理の流れを示したフローチャートである。文書入力部１０１は、電子ファイルを入力データとして読み込む（ステップＳ１００１）。電子ファイルは、モデム２０９からインターネット経由で読み込んでもよいし、場合によっては、既にハードディスク２０３に格納されているものを読み出してもよい。

【００７８】つぎに、文書種判定部１３１は、ファイルの先頭を読み出し、入力した文書がＸＭＬ文書ファイルであるか否かを判定する（ステップＳ１００２）。ＸＭＬ文書でないときは（ステップＳ１００２：ＮＯ）、処理を終了し、ＸＭＬ文書であるときは（ステップＳ１００２：ＹＥＳ）、文書型宣言と文書エンティティを読み込む（ステップＳ１００３）。要素分離部１３２はこの読み込まれた文書型宣言と文書エンティティとを、タグと要素の内容とに分離する（ステップＳ１００４）。

【００７９】要素解析部１３３は、分離されたタグと要素の内容から構文解析木を構築管理し、この際、図2に示した各タグの入籠構造を解析する。すなわち、要素解析部１３３は、木構造中の各要素の深さを求める（ステップＳ１００５）。たとえば、図2の例では、タグ「ｄｏｃ」を重要度ｌｅｖｅｌ１と判定し、１段階下のタグである「ｔｉｔｌｅ」、「ｍａｉｎ」、「ｍｉｓｃ」、「ｉｍｇ」は重要度ｌｅｖｅｌ２と判定し、以下同様に木構造の深さにしたがってｌｅｖｅｌ３、ｌｅｖｅｌ４を判定する。

【００８０】出力決定部１３４は、構文解析木により管理される要素のうち、出力すべき要素を、木構造の深さにしたがって決定する（ステップＳ１００６）。すなわ

15

ち、木構造が深ければそれだけその要素ないし下層に展開される要素が重要であると言えるので、出力決定部134は、木構造の深さにしたがって出力する要素を決定するのである。

【００８１】つづいて、文書整形部135は、スタイルシート240を読み込み、タグにしたがって、要素の内容を処理し、プリンタドライバを介して文書をプリントアウト可能に整形する（ステップS1007）。最後に、プリンタ206で整形された内容を印刷する（ステップS1008）。

【００８２】以上説明したように、実施の形態2のPCは、構造化文書の木構造の深さにしたがって要素の内容を出力するので、要素の物理量といった画一的な判断によらず、要素の意味内容を反映した出力が可能となる。また、伝達したい要素の内容にかかるタグについてはその構造が複雑になるので、すなわち、要素の階層が深くなる傾向があるので、構造化文書作成者が特に意識しなくても重要な要素の内容を閲覧者ないし利用者に伝達することができる。

【００８３】なお、実施の形態1ではタグの種類を、実施の形態2では要素の階層の深さを重要度の判定基準、すなわち、出力させるか否かを決定する基準としていたが、使用の態様によっては、図5に示したカテゴリないしタグの種類と、図2に示した要素の階層の深さとを両方用いて、出力すべき重要な要素を選択してもよい。この様にすることにより、重要度の判定精度をさらに向上させることができる。

【００８４】判定の手順としては、まず木構造の深さに着目して重要度を判定する。すなわち、重要度の高いと想定される要素を選択する。このとき、同一の深さのタグが複数ある場合に、図5に示したテーブルを参照し、重要度の高い要素を細分化して重要度の高い要素を出力する。

【００８５】たとえば、図2の構文解析木について、実施の形態2の方法で木構造の深さについて重要度を判定した場合、タグ「ｓｅｃｔｉｏｎｔｉｔｌｅ」、「ｃｏｎｔｅｎｔｓ」は共に重要度ｌｅｖｅｌ４となるが、図5のテーブルを参照すると「ｓｅｃｔｉｏｎｔｉｔｌｅ」は「ｃｏｎｔｅｎｔｓ」よりも重要度が高いタグとして登録されている。この場合はタグ「ｓｅｃｔｉｏｎｔｉｔｌｅ」の重要度はそのままｌｅｖｅｌ４とし、タグ「ｃｏｎｔｅｎｔｓ」の重要度をさらに一段階さげｌｅｖｅｌ５と判定する。この方法を用いれば実施例2と比較してさらに詳細な重要度の判定が可能となり精度が向上する。

【００８６】なお、本実施の形態1または2で説明した構造化文書の処理は、あらかじめ用意されたプログラムをパーソナル・コンピューターやワークステーション等のコンピュータで実行することにより実現することができる。すなわち、本発明は、コンピュータ上で単一のソフトウェア処理により実施することも可能である。このソフトウェア処理はコンピュータプログラムにより実現され、フロッピー（登録商標）ディスクやＣＤ－ＲＯＭ、ハードディスクなどの記録媒体に保存し、必要に応じてコンピュータに読み込み実行する。

16

【００８７】

【発明の効果】以上説明したように、本発明の構造化文書処理装置（請求項1）は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、構造化文書入力手段が、前記構造化文書記述言語により記述された構造化文書を入力し、出力決定手段が、前記構造化文書入力手段により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定し、構造化文書処理手段が、前記出力決定手段により出力させると決定された要素を出力可能に処理するので、タグを基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定でき、これにより、重要な要素を出力させる際の判定精度を高める構造化文書処理装置を提供することができる。

【００８８】また、本発明の構造化文書処理装置（請求項2）は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理装置であって、構造化文書入力手段が、前記構造化文書記述言語により記述された構造化文書を入力し、出力決定手段が、前記構造化文書入力手段により入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定し、構造化文書処理手段が、前記出力決定手段により出力させると決定された要素を出力可能に処理するので、タグと各要素間の論理構造を基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定でき、これにより、重要な要素を出力させる際の判定精度を高める構造化文書処理装置を提供することができる。

【００８９】また、本発明の構造化文書処理装置（請求項3）は、請求項1または2に記載の構造化文書処理装置において、構造化文書出力手段が、前記構造化文書処理手段により出力可能に処理された前記要素の内容を出力するので、構造化文書のうちの重要な要素の内容を出力でき、これにより、重要な要素を出力させる際の判定精度を高める構造化文書処理装置を提供することができる。

【００９０】また、本発明の構造化文書処理装置（請求項4）は、請求項3に記載の構造化文書処理装置において、前記構造化文書出力手段が、表示装置もしくは印刷装置であるので、重要な要素を出力させる際の判定精度を高める構造化文書処理装置を提供することができる。

【００９１】また、本発明の構造化文書処理方法（請求項5）は、構造化文書記述言語により記述された要素の

内容を出力可能に処理する構造化文書処理方法であって、構造化文書入力工程では、前記構造化文書記述言語により記述された構造化文書を入力し、出力決定工程では、前記構造化文書入力工程で入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定し、構造化文書処理工程では、前記出力決定工程で出力させると決定された要素を出力可能に処理し、構造化文書出力工程では、前記構造化文書処理工程で出力可能に処理された前記要素の内容を出力するので、タグを基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定でき、これにより、重要な要素を出力させる際の判定精度を高める構造化文書処理方法を提供することができる。

【００９２】また、本発明の構造化文書処理方法(請求項6)は、構造化文書記述言語により記述された要素の内容を出力可能に処理する構造化文書処理方法であって、構造化文書入力工程では、前記構造化文書記述言語により記述された構造化文書を入力し、出力決定工程では、前記構造化文書入力工程で入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定し、構造化文書処理工程では、前記出力決定工程で出力させると決定された要素を出力可能に処理し、構造化文書出力工程では、前記構造化文書処理工程で出力可能に処理された前記要素の内容を出力するので、タグと各要素間の論理構造を基に重要度を判定し、構造化文書内で重要な意味を持つ要素を特定でき、これにより、重要な要素を出力させる際の判定精度を高める構造化文書処理方法を提供することができる。

【００９３】また、本発明の構造化文書処理方法(請求項7)は、請求項5または6に記載の構造化文書処理方法において、前記構造化文書出力工程では、表示装置もしくは印刷装置を介して前記要素の内容を出力するので、重要な要素を出力させる際の判定精度を高める構造化文書処理方法を提供することができる。

【００９４】また、本発明のプログラム(請求項8)は、構造化文書記述言語により記述された要素の内容を出力可能に処理するプログラムであって、コンピュータを、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、前記構造化文書入力手段により入力された構造化文書中の各要素のタグの種類にしたがって当該要素を出力させるか否かを決定する出力決定手段と、前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段として機能させるので、タグを基に重要度を判定させ、構造化文書内で重要な意味を持つ要素を特定させることができ、これにより、重要な要素を出力させる際の判定精度を高めるプログラムを提供することができる。

【００９５】また、本発明のプログラム(請求項9)は、構造化文書記述言語により記述された要素の内容を出力可能に処理するプログラムであって、コンピュータを、前記構造化文書記述言語により記述された構造化文書を入力する構造化文書入力手段と、前記構造化文書入力手段により入力された構造化文書の要素中に定義されるタグの個数にしたがって、もしくは、当該要素中に定義されるタグの中でさらにタグが段階的に定義される場合の当該階層の深さにしたがって、当該要素を出力させるか否かを決定する出力決定手段と、前記出力決定手段により出力させると決定された要素を出力可能に処理する構造化文書処理手段、として機能させるので、タグと各要素間の論理構造を基に重要度を判定させ、構造化文書内で重要な意味を持つ要素を特定させることができ、これにより、重要な要素を出力させる際の判定精度を高めるプログラムを提供することができる。

【００９６】また、本発明のプログラム(請求項10)は、請求項8または9に記載のプログラムにおいて、さらに、コンピュータを、前記構造化文書処理手段により出力可能に処理された前記要素の内容を出力する構造化文書出力手段として機能させるプログラムを含んだので、構造化文書の重要な要素の内容を出力させることができ、これにより、重要な要素を出力させる際の判定精度を高めるプログラムを提供することができる。

【図面の簡単な説明】

【図1】実施の形態1で使用される、XMLが使用された構造化文書の構成例を示した図である。

【図2】構造化文書のタグの入籠構造の一例を示した説明図である。

【図3】実施の形態1の構造化文書処理装置をPCに適用した例の機能ブロック図である。

【図4】実施の形態1の構造化文書処理装置をPCに適用した場合の構成例を示した説明図である。

【図5】テーブル格納部に格納されたテーブルの内容の一例を表した説明図である。

【図6】テーブル格納部に格納されたテーブルのカテゴリと、タグと、重要度との関係の他の例を示した図である。

【図7】CSS(Cascading Style Sheets)と呼ばれるスタイルシートの一例を示した図である。

【図8】図1に示した構造化文書の出力例を示した図である。

【図9】実施の形態1の構造化文書処理装置をPCに適用した場合の構造化文書の処理の流れの一例を示したフローチャートである。

【図10】実施の形態2の構造化文書処理装置をPCに適用した場合の構造化文書の処理の流れを示したフローチャートである。

【図11】HTMLで記述した文書情報(ソース)の一

19

例を示した図である。

【図12】XMLで記述した文書情報(ソース)の一例を示した図である。

【図13】図11または図12で示したソースに基づいて構造化文書を出力した例である。

【符号の説明】

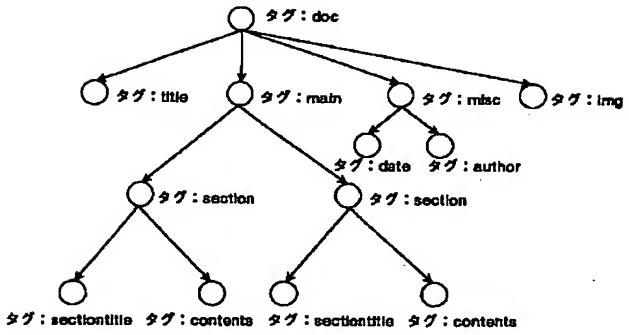
101　文書入力部
102　格納部
103　文書処理部
104　文書出力部
105　重要度指定部
121　文書格納部
122　テーブル格納部
123　スタイルシート 格納部
131　文書種判定部
132　要素分離部
133　要素解析部
134　出力決定部
135　文書整形部
203　ハードディスク
206　プリンタ
207　キーボード
209　モデム
232　XML解析アプリケーション
233　プリンタドライバ
234　ブラウザ
235　構造化文書ファイル
239　テーブル
240　スタイルシート

20

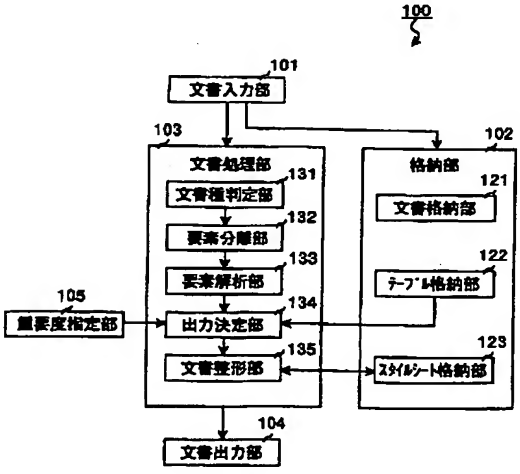
【図1】

```
1：<?xml version="1.0"encoding="Shift_JIS"?>
2：
3：<!ELEMENT doc (title,main,misc,img) >
4：<!ELEMENT title (#PCDATA) >
5：<!ELEMENT main (section*) >
6：<!ELEMENT section (sectiontitle,contents) >
7：<!ELEMENT sectiontitle (#PCDATA) >
8：<!ELEMENT contents (#PCDATA) >
9：<!ELEMENT misc (date,author) >
10：<!ELEMENT date (#PCDATA) >
11：<!ELEMENT author (#PCDATA) >
12：<!ELEMENT img EMPTY>
13：
14：<doc>
15：<title>タイトル</title>
16：<main>
17：<section>
18：<sectiontitle>1章のタイトル</sectiontitle>
19：<contents>1章の内容。</contents>
20：</section>
21：<section>
22：<sectiontitle>2章のタイトル</sectiontitle>
23：<contents>2章の内容。</contents>
24：</section>
25：</main>
26：<misc>
27：<date>99年8月1日</date>
28：<author>作成者名</author>
29：</misc>
30：<img src="imagefile.gif"> </img>
31：</doc>
```

【図2】



【図3】

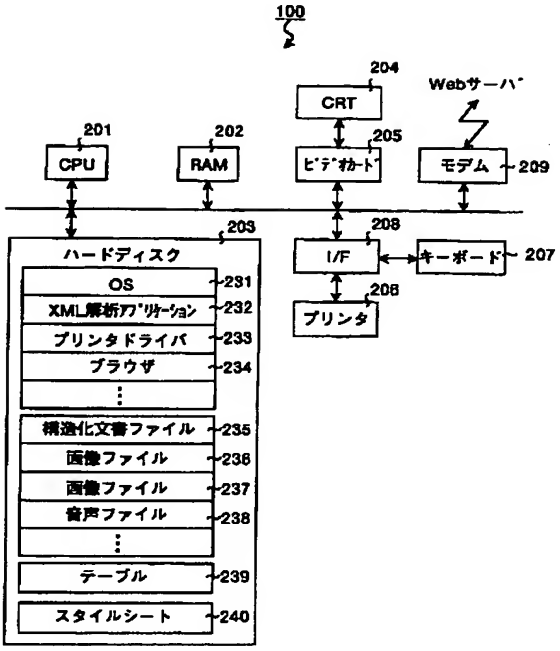


【図7】

```
title {
font-size :      32pt ;
font-weight :    bold ;
}

sectiontitle {
font-size :      18pt ;
font-style :          italic ;
```

【図４】



【図５】

239

カテゴリ	タグ	重要度
C1	title,Title, タイトル, …	level1
C2	sectiontitle, 章見出し, …	level2
C3	contents, Contents, img, …	level3
…….	…….	…….

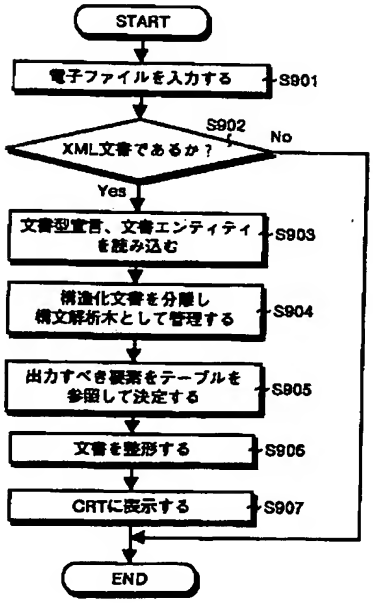
【図８】

タイトル
1章のタイトル
2章のタイトル

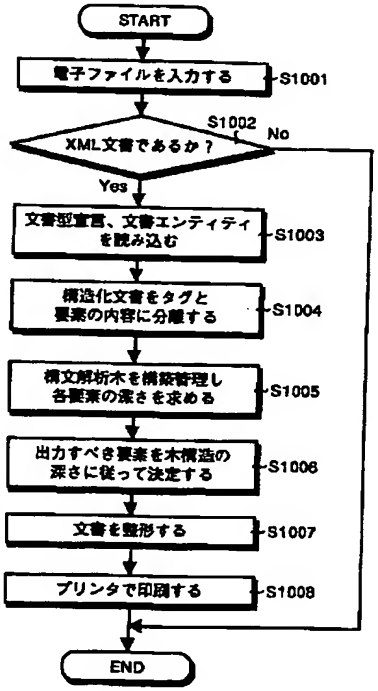
【図６】

カテゴリ	タグ	重要度
C1	title,Title	level1
	タイトル	level2
⋮	⋮	⋮

【図９】



【図１０】



【図１１】

```
<HTML>
<BODY>
<H1>タイトル</H1>
<H2>1章のタイトル</H2>
<P>1章の内容。</P>
<H2>2章のタイトル</H2>
<P>2章の内容。</P>
<H3>99年8月1日</H3>
<H3>作成者名</H3>
<IMG SRC="imagefile.gif">
</BODY>
</HTML>
```

【図１２】

```
<?xml version="1.0"encoding="Shift_JIS"?>

<!ELEMENT doc (title,main,misc,img) >
<!ELEMENT title (#PCDATA) >
<!ELEMENT main (section*) >
<!ELEMENT section (sectiontitle,contents) >
<!ELEMENT sectiontitle (#PCDATA) >
<!ELEMENT contents (#PCDATA) >
<!ELEMENT misc (date,author) >
<!ELEMENT date (#PCDATA) >
<!ELEMENT author (#PCDATA) >
<!ELEMENT img EMPTY>

<doc>
<title>タイトル</title>
<main>
<section>
<sectiontitle>1章のタイトル</sectiontitle>
<contents>1章の内容。</contents>
</section>
<section>
<sectiontitle>2章のタイトル</sectiontitle>
<contents>2章の内容。</contents>
</section>
</main>
<misc>
<date>99年8月1日</date>
<author>作成者名</author>
</misc>
<img src="imagefile.gif"> </img>
</doc>
```

【図１３】